PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Influence of serum iron test results on the diagnosis of iron deficiency in children: a retrospective observational study
AUTHORS	Sezgin, Gorkem; Li, Ling; Westbrook, Johanna; Wearne, Elisabeth; Azar, Denise; McLeod, Adam; Pearce, Christopher; Ignjatovic, Vera; Monagle, Paul; Georgiou, A

VERSION 1 – REVIEW

REVIEWER	Mantadakis, E Democritus University of Thrace, Pediatrics
REVIEW RETURNED	27-Dec-2020

GENERAL COMMENTS	The manuscript “Accurate diagnosis of iron deficiency in children: an observational study” is a retrospective observational study of children and adolescents 2 to 18 years of age followed in 137 general practices in Victoria, Australia. The authors using data derived from the electronic health records convincingly show that the diagnosis of iron deficiency is affected by the serum iron results, augmented by low hemoglobin test results and female sex. In other words, serum iron levels lead to cognitive errors in the correct interpretation of iron deficiency. The authors should be congratulated for this study, which is the first of its kind that quantitatively demonstrates the interpretation biases of serum iron in diagnosing iron deficiency. Research focusing on diagnostic errors in Pediatrics is very limited and incorrect interpretation of common diagnostic tests can lead to adverse clinical outcomes. By using appropriate definitions, the authors showed that children without iron deficiency (serum ferritin >30μg/dL) but with a low serum iron level were 4 times more likely to be erroneously diagnosed with iron deficiency (overdiagnosis). On the other hand, children with true iron deficiency based on a low serum ferritin <12μg/dL were 0.7 times less likely to be correctly diagnosed based on a normal serum iron level (underdiagnosis). The study is obviously limited by its retrospective observational design and the limited data recorded in the electronic health software, something that the authors correctly acknowledge towards the end of the Discussion section of the manuscript. On the other hand, their sample was very large and likely representative of the patients tested for iron deficiency. Moreover, they used different ferritin thresholds to define their iron deplete and iron replete populations. Major comment: The study is well done but I disagree with the exclusion of toddlers 1-2 years of age. In fact, iron deficiency and iron deficiency anemia are much more common in toddlers versus older children, and the debilitating cognitive effects of iron deficiency are mainly limited to this age group that was not studied at all. This is a serious limitation of the study. They authors should provide

	convincing evidence, i.e., relative references for their statement about the uncertain diagnostic criteria for iron deficiency in toddlers, which I believe is an overstatement. Also, this study should be followed by a prospective interventional study on the same general practices after relabeling of the diagnostic studies for iron deficiency, so that standalone ferritin be the requested study by general practitioners in diagnosing iron deficiency. Although I am not familiar with the medical system in Australia, it would be nice to see whether Pediatricians and not General Practitioners are also predisposed to the same cognitive errors introduced by including serum iron in the diagnostic panel of iron deficiency. Minor comments:1)The reported value of CRP (page 8, line 15) is likely wrong, i.e., it is 10mg/dL and not 10mg/ml. Please, clarify. 2)In page 12, line 59 (last line of page 12), the sentence: Alternatively, it may a result…. Should be corrected to: Alternatively, it may be a result…

REVIEWER	Renella, Raffaele Lausanne University Hospital
REVIEW RETURNED	30-Dec-2020

GENERAL COMMENTS	Sezgin and colleagues report of a large retrospective observational study of n>13k children aged 2-8 y.o. who were investigated in 137 GP office settings in the most densely populated area of Australia for iron deficiency over a period of 10 years. The aim of the study was to question the association of serum iron levels with the diagnosis of ID by GPs in two groups of children (ID or no ID by serum ferritin levels). This study was performed according to research ethics regulations, after approval of the local IRB, and involved chart reviews with subsequent de-identification (no direct patient involvement). The chosen methodology is appropriate and clearly outlined in the manuscript. Inclusion and exclusion criteria, as well as definitions of variables and categories, are clearly outlined and are globally appropriate. The statistical analysis appears sound and appropriate, although this latter aspect is outside my specific professional expertise. The authors need to be congratulated for asking the important and clinically relevant question of whether an appropriate testing strategy for ID in children by GPs should include the measurement of serum iron levels. The appropriate diagnosis of ID in children is a major healthcare priority. The authors provide highly interesting evidence to support the notion that adjusted frontline lab testing by GPs should possibly not include serum iron levels, or that interventions are required to educate the GPs on its value for the diagnosis of ID in children (or absence thereof). Their study shows that the inclusion of serum iron levels in lab panels causes confusion and misdiagnosis of ID in children by GPs (i.e., both the underdiagnosis in iron deplete children and the overdiagnosis in iron replete children). Furthermore, the data shown in this study suggests that biases (availability/non availability and information distortion) might influence the ID diagnosis, as anemia and female sex are associate with an incorrect ID diagnosis in iron replete children as well. The manuscript is well written, informative and educative. The main limitations of the study are associated to the data source (retrospective electronic chart review), and they are discussed

	partially. A detailed review is at this time hindered by the unavailability of the complete/raw set of data of the study. The tables and the supplementary document provide a summary of the findings and the statistical assessments, but there is no file including all the numerical variables which contributed to the results. For example, and amongst many others, it would be interesting to know what proportion of the tests were associated with the clear mention of "ID/IDA diagnosis" in the chart. The availability of such data would strengthen the authors claims. Other concerns: 1. The title should be more precise and refer to the main findings of the study. In its current form it will fail to attract the reader to the important data of the study. 2. One of the main weaknesses of the study is the absence of information on the reasons/motivation behind performing the initial iron testing in children in the study. This information might in fact be the main modulator of the decision/diagnosis-making by GPs. If a 14 y.o. female with fatigue, asthenia or a 3 y.o. with pallor presents to the GP office, the diagnosis of ID/provision of iron supplements might be driven by clinical data, invisible to the investigators. I understand the investigators attempt to approach this issue via the inclusion of proxies such as anemia and sex, but I think it would be important to discuss this limitation resulting from the study design rather from GP biases. How can the authors be certain that the GP was truly influenced by the serum iron results in the diagnosis of ID and not the upfront motive for consultation or other clinical elements? 3. How many children were tested multiple times, and treated multiple times? If there were 17k test pairs (serum ferritin-serum iron) for 13k children, this means there were duplicate tests? Had the test pairs (serum ferritin-serum iron) to be performed the same day? 4. The administration of iron supplements can sometimes be initiated by parents/families without the requirement of a healthcare provider prescription. Is this the case in the setting of the study? How likely is the notion of "treatment of ID/IDA" in the electronic chart to realistically constitute the proxy for the GP diagnosis of ID/IDA? Could the GP record supplementation of iron given by parents without him formally making the diagnosis of ID/IDA or prescribing such supplement? In the study design this would cause the incorrect attribution of the diagnosis of ID/IDA to that child. 5. It is not clear whether any of the children received a transfusion (pRBCs) prior to the inclusion into the study. These patients should possibly be excluded (any lifetime transfusion of pRBC) as they could suggest co-morbidities/confounders. 6. Do the authors believe patients with IBDs or chronic renal disorders (CKDs) should be excluded from the study? Children with IBDs can be iron deplete by have "falsely" increased ferritin levels and close to normal acute reactants (i.e. CRP levels). These diagnoses should be available from the electronic charts of the patients. 7. Did any of the patients that were iron replete and had high serum iron levels receive the diagnosis of iron overload disorders? If the GP performed the iron lab testing for another reason than the diagnosis of ID, were these cases excluded from the dataset? This point also links to point 2 above.

	8. The authors should (in the discussion) provide their views on why serum iron levels are frequently used by GPs in Australia. Is this because of the role of serum iron levels in the calculation of transferrin saturation levels? Does their local healthcare authority recommend a “standard iron panel” in children? If so, does it include serum iron levels and why? Do they have any other hypothesis? 9. The authors should provide a counterpoint to their argument highlighting the value of a complete iron panel and the value of its appropriate interpretation. The notion that a test should not be performed because it is confusing for the physician prescribing/interpreting it, is debatable. Therefore, the authors should provide their perspectives on how they believe the confusion/misdiagnoses of ID caused by the dosage of serum levels should be resolved? Do they suggest recommendations should specify that upfront iron testing for ID should only be based on ferritin levels? Should there be more education of GPs on the interpretation of lab panels including serum iron levels? Are there any local standards for the diagnosis of ID in children, and how should they be modified? 10. The authors should provide elements to support the applicability of their observations to other settings. Do they believe their findings are universal? Would their findings be replicated in environments with higher prevalence of hemoglobinopathies, with lower resources?

REVIEWER	Gichohi-Wainaina, Wanjiku ICRISAT
REVIEW RETURNED	30-Dec-2020

GENERAL COMMENTS	-The author does explain that often due to inflammation, serum ferritin results become complicated to interpret and therefore serum iron results may be more useful.However, this does not come out clearly in the introduction....a bit more could have been explained on why physicians prefer to use serum iron results to interpret iron deficiency -Reference 14 should be updated-WHO has an updated guideline for the assessment of iron deficiency, though the cut offs remain largely the same, the literature should reflect the latest cut offs -Had the patients been informed prior that their records may be used for research purposes? -The population was between 2-18 years but to denote iron depletion only a cut off of 12 was used..which is suitable for those below 5 years but for 5 and above years a cut off of 15 needed to be used -In the methods section-the cut offs for normal serum iron should have been defined as well -How did the authors decide which co-variates to included in the GEE? - The author rightly states in the abstract thats serum iron should not be used for diagnosis of Fe deficiency...WHO guidelines are also clear.The publication is not convincing on the utility of this work for clinical practice or indeed public health.Would it not be easier to train physicians/pathologists on the correct measures of diagnoses are?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

The manuscript "Accurate diagnosis of iron deficiency in children: an observational study" is a retrospective observational study of children and adolescents 2 to 18 years of age followed in 137 general practices in Victoria, Australia. The authors using data derived from the electronic health records convincingly show that the diagnosis of iron deficiency is affected by the serum iron results, augmented by low hemoglobin test results and female sex. In other words, serum iron levels lead to cognitive errors in the correct interpretation of iron deficiency.

The authors should be congratulated for this study, which is the first of its kind that quantitatively demonstrates the interpretation biases of serum iron in diagnosing iron deficiency. Research focusing on diagnostic errors in Pediatrics is very limited and incorrect interpretation of common diagnostic tests can lead to adverse clinical outcomes. By using appropriate definitions, the authors showed that children without iron deficiency (serum ferritin >30μg/dL) but with a low serum iron level were 4 times more likely to be erroneously diagnosed with iron deficiency (overdiagnosis). On the other hand, children with true iron deficiency based on a low serum ferritin <12μg/dL were 0.7 times less likely to be correctly diagnosed based on a normal serum iron level (underdiagnosis).

The study is obviously limited by its retrospective observational design and the limited data recorded in the electronic health software, something that the authors correctly acknowledge towards the end of the Discussion section of the manuscript. On the other hand, their sample was very large and likely representative of the patients tested for iron deficiency. Moreover, they used different ferritin thresholds to define their iron deplete and iron replete populations.

On behalf of my co-authors, I would like to thank the reviewer for their comments and their review of our study.

Major comment: The study is well done but I disagree with the exclusion of toddlers 1-2 years of age. In fact, iron deficiency and iron deficiency anemia are much more common in toddlers versus older children, and the debilitating cognitive effects of iron deficiency are mainly limited to this age group that was not studied at all. This is a serious limitation of the study. They authors should provide convincing evidence, i.e., relative references for their statement about the uncertain diagnostic criteria for iron deficiency in toddlers, which I believe is an overstatement.

Children aged one years of age have now been included in the analysis. Two year olds were already included. The results have also been updated to reflect this change.

The statement regarding uncertain diagnostic criteria has now been removed.

Also, this study should be followed by a prospective interventional study on the same general practices after relabeling of the diagnostic studies for iron deficiency, so that standalone ferritin be the requested study by general practitioners in diagnosing iron deficiency. Although I am not familiar with the medical system in Australia, it would be nice to see whether Pediatricians and not General Practitioners are also predisposed to the same cognitive errors introduced by including serum iron in the diagnostic panel of iron deficiency.

We thank the reviewer for their comment, however, such an intervention is beyond the scope of this study. As indicated in our discussion, recommendations were made by authoritative bodies to relabel iron panel test with clinically relevant terms so that clinicians would be encouraged to order and thus interpret correct tests, across Australia. A future study, beyond our scope, could be conducted to measure the impact of such an intervention, if implemented.

1)The reported value of CRP (page 8, line 15) is likely wrong, i.e., it is 10mg/dL and not 10mg/ml. Please, clarify.

We thank the reviewer for noticing this error. The unit has now been corrected to mg/L. Please note that we have verified that it is 'L' in our dataset, rather than 'dL', and the indicator for inflammatory disease is indeed >10 mg/L.

2)In page 12, line 59 (last line of page 12), the sentence: Alternatively, it may a result…. Should be corrected to: Alternatively, it may be a result…

The correction has now been made.

Reviewer: 2

Sezgin and colleagues report of a large retrospective observational study of n>13k children aged 2-8 y.o. who were investigated in 137 GP office settings in the most densely populated area of Australia for iron deficiency over a period of 10 years.

The aim of the study was to question the association of serum iron levels with the diagnosis of ID by GPs in two groups of children (ID or no ID by serum ferritin levels).

This study was performed according to research ethics regulations, after approval of the local IRB, and involved chart reviews with subsequent de-identification (no direct patient involvement). The chosen methodology is appropriate and clearly outlined in the manuscript. Inclusion and exclusion criteria, as well as definitions of variables and categories, are clearly outlined and are globally appropriate. The statistical analysis appears sound and appropriate, although this latter aspect is outside my specific professional expertise.

The authors need to be congratulated for asking the important and clinically relevant question of whether an appropriate testing strategy for ID in children by GPs should include the measurement of serum iron levels. The appropriate diagnosis of ID in children is a major healthcare priority. The authors provide highly interesting evidence to support the notion that adjusted frontline lab testing by GPs should possibly not include serum iron levels, or that interventions are required to educate the GPs on its value for the diagnosis of ID in children (or absence thereof). Their study shows that the inclusion of serum iron levels in lab panels causes confusion and misdiagnosis of ID in children by GPs (i.e., both the underdiagnosis in iron deplete children and the overdiagnosis in iron replete children). Furthermore, the data shown in this study suggests that biases (availability/non availability and information distortion) might influence the ID diagnosis, as anemia and female sex are associate with an incorrect ID diagnosis in iron replete children as well.

The manuscript is well written, informative and educative. The main limitations of the study are associated to the data source (retrospective electronic chart review), and they are discussed partially.

On behalf of my co-authors, I would like to thank the reviewer for their comments and their review of our study.

A detailed review is at this time hindered by the unavailability of the complete/raw set of data of the study. The tables and the supplementary document provide a summary of the findings and the statistical assessments, but there is no file including all the numerical variables which contributed to the results. For example, and amongst many others, it would be interesting to know what proportion of the tests were associated with the clear mention of "ID/IDA diagnosis" in the chart. The availability of such data would strengthen the authors claims.

We are unfortunately unable to provide raw dataset due to data use and ethics agreements, which is common with similar datasets and sources in Australia. Information regarding data access is detailed in our data sharing statement. A study on the dataset used in our analysis, which may provide a better understanding of the variables, by Pearce et al. is referenced in our study (reference 10). Limitations/strengths of our outcome variables and the dataset are discussed in the discussion section of our manuscript.

1. The title should be more precise and refer to the main findings of the study. In its current form it will fail to attract the reader to the important data of the study.

The title has now been updated to improve precision, taking into account the guidelines of BMJ Open on title requirements. It now reads:

"Influence of serum iron test results on the diagnosis of iron deficiency in children: a retrospective observational study."

2. One of the main weaknesses of the study is the absence of information on the reasons/motivation behind performing the initial iron testing in children in the study. This information might in fact be the main modulator of the decision/diagnosis-making by GPs. If a 14 y.o. female with fatigue, asthenia or a 3 y.o. with pallor presents to the GP office, the diagnosis of ID/provision of iron supplements might be driven by clinical data, invisible to the investigators. I understand the investigators attempt to approach this issue via the inclusion of proxies such as anemia and sex, but I think it would be important to discuss this limitation resulting from the study design rather from GP biases. How can the authors be certain that the GP was truly influenced by the serum iron results in the diagnosis of ID and not the upfront motive for consultation or other clinical elements?

Regardless of this limitation, diagnosis of iron deficiency can only be confirmed with laboratory testing (serum ferritin). Any clinical sign leading to a suspicion of iron deficiency, including non-specific symptoms such as fatigue or asthenia should be followed up with a laboratory test. Diagnosis of a patient with iron deficiency without laboratory confirmation is not within guidelines.

We have now added the uncertainties about reasons for test orders in the limitations section of discussion, as such:

"The reason for ordering of iron panel tests is not known, and may have an impact on the diagnosis. Nonetheless, serum ferritin is a test used in checking the iron status of a patient, and low levels, particularly < 12/15 ug/L should be treated."

3. How many children were tested multiple times, and treated multiple times? If there were 17k test pairs (serum ferritin-serum iron) for 13k children, this means there were duplicate tests? Had the test pairs (serum ferritin-serum iron) to be performed the same day?

Duplicate tests are indeed a possibility in the dataset, and has been addressed in our study inclusion/exclusion criteria:

"To ensure our study population contained only new cases, we excluded cases where a serum ferritin or serum iron test was ordered within the last 12 months".

This includes possible same day tests. Therefore the duplicate tests are likely independent of a previous test.

4. The administration of iron supplements can sometimes be initiated by parents/families without the requirement of a healthcare provider prescription. Is this the case in the setting of the study? How likely is the notion of "treatment of ID/IDA" in the electronic chart to realistically constitute the proxy for the GP diagnosis of ID/IDA? Could the GP record supplementation of iron given by parents without him formally making the diagnosis of ID/IDA or prescribing such supplement? In the study design this would cause the incorrect attribution of the diagnosis of ID/IDA to that child.

Iron medications can indeed be obtained without a prescription, although the dosage required to treat children with serum ferritin levels <12/15 ug/L is unlikely to be achieved with such medications. This is indicated in the limitations section of the discussion:

"Although medication prescriptions are well recorded, iron deficiency medication can be obtained without prescription; however the dosage of over the counter iron supplements may not be appropriate for patients with <12/15 ug/L serum ferritin, particularly for paediatric patients."

Our medications data is from medication prescriptions given out by the general practitioner, and are therefore are not expected to be records of medications given by other parties. We have made the following adjustment in the methods section for clarity:

"Medications used in the treatment of iron deficiency were identified by searching for commercial and generic names of iron supplements in the field record of medications prescribed by the general practitioner."

5. It is not clear whether any of the children received a transfusion (pRBCs) prior to the inclusion into the study. These patients should possibly be excluded (any lifetime transfusion of pRBC) as they could suggest co-morbidities/confounders.

6. Do the authors believe patients with IBDs or chronic renal disorders (CKDs) should be excluded from the study? Children with IBDs can be iron deplete by have "falsely" increased ferritin levels and close to normal acute reactants (i.e. CRP levels). These diagnoses should be available from the electronic charts of the patients.

We would like to address these two points together, as they are of similar nature. As the reviewer has correctly noted, there are certain factors which may confound the interpretation of serum ferritin levels, including pRBCs and CKDs. We have addressed some of the major confounders, including inflammation (through CRP tests), previous serum ferritin tests, past medications and diagnoses related to iron deficiency, and to an extent, thalassemia. Nonetheless, there are countless other potential confounders, and many other factors which come into play when a diagnosis is being made with any diagnostic procedure. Chronic diseases of patients and clinical investigations are not always recorded in general practice and we cannot reliably exclude all potential confounders. We have now added the following phrase into the limitations section to discuss this limitation, as such:

"Certain diseases may impact the interpretation of serum ferritin and the diagnosis of iron deficiency, including thalassemia and inflammatory diseases. We accounted for a number of these potential confounders, although it is likely that other unknown factors (diseases, clinical investigations, patient history) may have impacted the interpretation of test results and the diagnosis of iron deficiency."

Nonetheless, we believe the results of our study are accurate even in the presence of such potential confounders. We present the effect of serum iron on the diagnosis of iron deficiency in two independent circumstances: iron replete and iron deplete, both in which we see the influence of serum iron.

7. Did any of the patients that were iron replete and had high serum iron levels receive the diagnosis of iron overload disorders? If the GP performed the iron lab testing for another reason than the diagnosis of ID, were these cases excluded from the dataset? This point also links to point 2 above.

Iron overload is beyond the scope of this study, as the focus of our study is iron deficiency. There may indeed have been patients diagnosed with iron overload.

Regardless of the reason for testing, a serum ferritin level of <12 ug/L indicates that the patient requires treatment for iron deficiency, and follow up. Furthermore, in the absence of inflammatory disease, a patient with serum ferritin levels >30 ug/L is not indicative of iron deficiency (we also include >50/100 ug/L in our supplementary document). We found that low serum iron increases the likelihood of an iron deficiency diagnosis in such cases.

8. The authors should (in the discussion) provide their views on why serum iron levels are frequently used by GPs in Australia. Is this because of the role of serum iron levels in the calculation of transferrin saturation levels? Does their local healthcare authority recommend a "standard iron panel" in children? If so, does it include serum iron levels and why? Do they have any other hypothesis?

As indicated in the introduction, serum ferritin is often not ordered standalone: an 'iron panel' test is ordered. This panel of test contains serum ferritin, serum iron, transferrin saturation, and transferrin.

We have also, in the discussion, indicated that other studies have found iron panel test being ordered much more than standalone ferritin tests in general practice.

We have discussed our views on why serum iron is being used instead of serum iron in the discussion:

"Our results suggest a misconception of the changed role of iron levels in diagnosing iron deficiency; serum iron being confused as the biomarker of iron deficiency where in fact, serum ferritin is. Serum iron performed as part of a panel act as a distractor in interpreting the test."

We therefore believe that serum iron is being interpreted because it is being confused as the test for diagnosing iron deficiency. We have discussed errors in diagnosis in our discussion, indicating that incorrect interpretation of test results is indeed not uncommon.

9. The authors should provide a counterpoint to their argument highlighting the value of a complete iron panel and the value of its appropriate interpretation. The notion that a test should not be performed because it is confusing for the physician prescribing/interpreting it, is debatable. Therefore, the authors should provide their perspectives on how they believe the confusion/misdiagnoses of ID caused by the dosage of serum levels should be resolved? Do they suggest recommendations should specify that upfront iron testing for ID should only be based on ferritin levels? Should there be more education of GPs on the interpretation of lab panels including serum iron levels? Are there any local standards for the diagnosis of ID in children, and how should they be modified?

We have now added a section discussing the value of serum iron in the introduction, as such:

"The utility of serum iron in clinical diagnosis is limited, particularly in general practice. Serum iron and transferrin saturation levels, when interpreted together, provide an indication of iron use by tissue. However, their use in diagnosing iron deficiency is not supported by current guidelines".

In the discussion, we have the following section which provides our perspective on how this issue may be addressed, backed by other studies:

"The Medicare taskforce made recommendations to relabel tests in clinically relevant terms, such that a clinician would order 'iron deficiency studies' containing ferritin, rather than the current iron panel test with all iron biomarkers. Such an approach may reduce cognitive errors by removing confounding factors out of the decision-making process."

10. The authors should provide elements to support the applicability of their observations to other settings. Do they believe their findings are universal? Would their findings be replicated in environments with higher prevalence of hemoglobinopathies, with lower resources?

We have the following section in our discussion:

"Although our results from Australian general practices within the states of Victoria, it is possible that a similar diagnostic error is occurring in other states/territories and countries"

We have modified this phrase as such, to improve clarity:

"…, it is possible that a similar diagnostic error is occurring in other states/territories and countries where iron panel tests are ordered in the investigation of iron deficiency instead of standalone serum ferritin.

<u>Reviewer: 3</u>

On behalf of my co-authors, I would like to thank the reviewer for their comments and their review of our study.

-The author does explain that often due to inflammation, serum ferritin results become complicated to interpret and therefore serum iron results may be more useful.However, this does not come out clearly in the introduction....a bit more could have been explained on why physicians prefer to use serum iron results to interpret iron deficiency

Due to inflammation, interpretation of serum ferritin does indeed become complicated. However, we do not say that serum iron results may be more useful in such cases, as this is not correct.

We have moved the section on CRP testing to methods (inclusion/exclusion criteria) to prevent confusion.

-Reference 14 should be updated-WHO has an updated guideline for the assessment of iron deficiency, though the cut offs remain largely the same, the literature should reflect the latest cut offs

Reference has now been updated.

Had the patients been informed prior that their records may be used for research purposes?

As per ethics guidelines, our research data has been de-identified and is non-identifiable, and has been collected with information about their data collection, with the option to opt-out given them by their general practitioners.

Our dataset has been collected in accordance with ethics guidelines, and has been approved, by an ethics committee to collect the data (by our data provider) and by another ethics committee to use in our research. This information is provided in the methods section of our manuscript. Furthermore, the dataset has been subject to a privacy impact assessment in accordance with Australian legislative requirements and Australian Privacy Principles.

-The population was between 2-18 years but to denote iron depletion only a cut off of 12 was used..which is suitable for those below 5 years but for 5 and above years a cut off of 15 needed to be used

Our study has now been adjusted to account for the different threshold for >5 years. These changes have been reflected in the results.

-In the methods section-the cut offs for normal serum iron should have been defined as well

There is no cut-off for serum iron, as it is not associated with iron deficiency or disease. The cut off for serum ferritin is relevant to iron deficiency, the cut-off for CRP is relevant to inflammatory disease, and the cut-off for haemoglobin is relevant to anaemia. We used the abnormality levels denoted by laboratories for serum iron, which may vary depending on the laboratory as reference intervals often differ by population.

-How did the authors decide which co-variates to included in the GEE?

We have added the following phrase to the methods section (Statistical analysis) to clarify the decision process for GEE:

"Our covariates were selected due to their relevance to the diagnosis of iron deficiency. Female sex and younger age have high prevalence of iron deficiency. Red cell indices are relevant to the diagnosis of iron deficiency anaemia, and serum transferrin is ordered alongside serum ferritin and can be an indicator of tissue iron absorption (if interpreted in consideration of serum iron levels), although its use in the diagnosis of iron deficiency is not recommended."

- The author rightly states in the abstract thats serum iron should not be used for diagnosis of Fe deficiency...WHO guidelines are also clear.The publication is not convincing on the utility of this work for clinical practice or indeed public health.Would it not be easier to train physicians/pathologists on the correct measures of diagnoses are?

In our discussion section, we have the following phrase indicating the significance of our study:

"Our results provide empirical evidence demonstrating the resultant diagnostic errors due to serum iron tests among children."

We contextualise this statement with previous findings from other studies, indicating that serum iron is commonly ordered in the diagnosis of iron deficiency.

Together, our study and this previous study suggest that serum iron test results are detrimental in the diagnosis of iron deficiency. Given that serum iron tests results are generally not of clinical significance in general practice, their removal from the diagnosis of iron deficiency could be beneficial (for clinical practice and public health), which we indicate as such:

"The Medicare taskforce made recommendations to relabel tests in clinically relevant terms, such that a clinician would order 'iron deficiency studies' containing ferritin, rather than the

current iron panel test with all biomarkers. Such an approach may reduce cognitive errors by removing confounding factors out of the decision making process."

We have now expanded on this by the inclusion of the following phrase:

"In the diagnosis of iron deficiency, this may reduce burden of disease by ensuring patients with iron efficiency are correctly and timely diagnosed, potentially preventing anaemia and cognitive impairment. Additionally, correct use of serum ferritin to diagnose iron deficiency may reduce the number of patients being misclassified as having iron defiance based on serum iron results, potentially reducing unnecessary treatment, and promoting correct diagnosis by ruling out iron deficiency."

Current training practices (e.g. textbooks) are consistent with guidelines, and recommend the use of serum ferritin in diagnosing iron deficiency.

VERSION 2 – REVIEW

REVIEWER	Mantadakis, E Democritus University of Thrace, Pediatrics
REVIEW RETURNED	19-Mar-2021

GENERAL COMMENTS	The authors have for the most part replied adequately to the criticism they received. The manuscript deserves publication.

REVIEWER	Renella, Raffaele Lausanne University Hospital
REVIEW RETURNED	26-Mar-2021

GENERAL COMMENTS	The majority of the comments in my prior review of this manuscript have been taken into account, and the main issues have been resolved. I have no further comments. Congratulations to the authors for their interesting manuscript.